IMAGE COMPRESSION USING A COLOR VISUAL MODEL

BACKGROUND OF THE INVENTION

This application claims the benefit of 60/441,583 filed January 21, 2003 entitled

5   Automatic Image Compression Using A Color Visual Model.

The present invention relates to a system for coding images, and more particularly, to a system for compressing images to a reduced number of bits by employing a Discrete Cosine Transform (DCT) in combination with a visual model.

There has been significant development in the compression of digital information for

10   digital images. The effective compression of digital information is important to maintain sufficient quality of the digital image while at the same time reducing the amount of data required for representing the digital image. The transmission of the digital images has gained particular importance in television systems and Internet based transmission. If the digital images include a relatively large number of bits to represent the digital images, a significant burden is

15   placed on the infrastructure of communication networks involved with the creation, transmission, and re-creation of digital images. For this reason, there is a need to compress digital images to a smaller number of bits, by reducing redundancy and "invisible" image components of the images themselves.

Still image compression techniques, such as JPEG, compress digital information for

20   digital images. As in digital compression for the transmission of digital video, JPEG compression includes a tradeoff between file size and compressed image quality. For example, JPEG compression is extensively used in digital cameras, Internet based applications, and databases containing digital images.

Many of the image compression techniques, such as JPEG amd MPEG, include a

25   transform coding algorithm for the digital image, wherein the image is divided into blocks of pixels. For example, each block of pixels may be an 8x8 or 16x16 block of pixels. Each block of pixels then undergoes a two dimensional transform to produce a two dimensional array of transform coefficients. For many image coding applications, a Discrete Cosine Transform (DCT) is utilized to provide an orthogonal transform. After the block of pixels undergoes a Discrete

30   Cosine Transform (DCT), the resulting transform coefficients are subject to compression by

<u>1</u>

thresholding and quantization operations. Thresholding involves setting all coefficients whose magnitude is smaller than a threshold value equal to zero, whereas quantization involves scaling a coefficient by step size and rounding off to the nearest integer.

Commonly, the quantization of each DCT coefficient is determined by an entry in a quantization matrix (Q-table). A quantization matrix includes a plurality of values that is used to group a set of values together. For example, a quantization matrix may be used to group the values from 0 to 3 into group 1, values from 3-6 into group 2, and values from 6-9 into group 3. It is this matrix that is primarily responsible for the perceived image quality and the bit rate of the transmission of the image. The perceived image quality is important because the human visual system can only tolerate a certain amount of image degradation without significantly observing a noticeable error. Therefore, certain images can tolerate significant degregration and thus be significantly compressed, whereas other images cannot tolerate significant degradation and should not be significantly compressed.

Some systems include computing a single DCT quantization matrix based on human sensitivity. One such system is based on a mathematical formula for the human contrast sensitivity function, scaled for viewing distance and display resolution, as taught in U.S. Patent No. 4,780,716. Another such system is based on a formula for the visibility of individual DCT basic functions, as a function of viewing distance, display resolution, and display luminance. The formula is disclosed in both a first article entitled "Luminance-Model-Based DCT Quantization For Color Image Compression" of A.J. Ahumada et al. published in 1992 in the Human Vision, Visual Processing, and Digital Display III Proc. SPIE 1666, Paper 32, and a second technical article entitled "An Improved Detection Model for DCT Coefficient Quantization" of H.A. Peterson, et al., published in 1993, in Human Vision, Visual Processing and Digital Display VI Proc. SPIE. Vol. 1913 pages 191-201. The techniques described in the '761 patent and the two technical articles do not adapt the quantization matrix to the image being compressed.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram of a computer network that may be used in the practice of the present invention.

FIG. 2 schematically illustrates a block diagram of an image encoding system.

FIG. 3 schematically illustrates the comparison of a pair of images.

DETAILED DESCRIPTION OF THE INVENTION

5   Referring to FIG. 1 a block diagram of a computer network 10 for the storing, retrieving, and transmitting of images is illustrated. A pair of image processing devices 12 and 14 are provided. The image processing device 12 may be used to perform a storage mode 16 and a retrieval mode 18 operation of the network 10 and, similarly, the image processing device 14 may be used to perform a storage mode 16 and a retrieval mode 18 operation of the network 10.

10   The storage mode 16 accesses a disk subsystem 20, whereas the retrieval mode 18 recovers information from the disk subsystem 20.. Each of the devices 12 and 14 may be any type of processing device, or otherwise a single processing device including the functionality of both devices 12 and 14. The devices 12 and 14 may further include a RAM 26, a communication channel 22, a CPU processor 24, and a display subsystem 28.

15   In general the system may include, in part, a compression technique that incorporates a Discrete Cosine Transform (DCT). In the storage mode 16, an image 30 including a plurality of pixels, represented by a plurality of digital bits, is received from any suitable sources through the communication channel 22 of the device 12. The device, and in particular the CPU processor 24, performs a DCT transformation, computes a DCT mask, if desired, selects a quantization

20   matrix, and estimates a quantization matrix optimizer. The device 12 then quantizes the digital bits comprising the image 30, and performs encoding of the resulting quantized DCT coefficients, such as by example by run-length encoding, Huffman coding, or arithmetic coding. The resulting quantization matrix is then stored in coded form along with coded coefficient data using any suitable technique, such as the JPEG standard. The compressed file is then stored on

25   the disk subsystem 20 of the device 12, or otherwise transmitted to another device.

          In the retrieval mode 18, the device 12 (or 14) retrieves the compressed file from the disk subsystem 20, and decodes the quantization matrix and the DCT coefficient data. The device 12 (or 14) then de-quantizes the coefficients by multiplication of the resulting scaled quantization matrix and performs an inverse DCT. The resulting digital file containing pixel data is available

30   for display on the display subsystem 28 of the device 12 (or 14) or can be transmitted to the

device 14 (or 12) or elsewhere by the communication channel 22. The resulting digital file is illustrated in FIG. 1 as 30' (IMAGE).

In some applications, such as digital image database applications, the image may be compressed using a Q-table and then the resulting compressed image is reconstructed and presented to the user. The user then makes adjustments to the Q-table in some fashion and the process is repeated until an acceptable compression of the image is achieved. While this achieves an acceptable result, the process is time consuming, especially for large digital image databases. While it is the case that the appropriate selection of a Q-table (set of values) is desirable, it is problematic to automatically select such a table.

One existing technique for the selection of the Q-table is illustrated in U.S. Patent No. 5,426,512, incorporated by reference herein. The error resulting from quantization for a given scale factor of the Q-table is scaled in the DCT domain by using a perceptual mask, that suppresses some errors and leaves some other errors. The result after applying the mask is then spatially pooled and compared against a target error. If sufficiently close to a target error, then the current Q-table is used to compress the image. If not sufficiently close, the Q-table is adjusted. The model used is based upon a mean block luminance (for light adaptation) and a DCT coefficient that depends on thresholds based on coefficient amplitudes (for masking).

After consideration of using a visual model within the compression process for Q-table optimization and comparison of DCT coefficients of compressed and uncompressed images, as disclosed in the '512 patent, the present inventors determined that the resulting model does not accurately reflect the user's perception of the images. Moreover, using the visual model within the compression process for Q-table optimization and comparison of DCT coefficients of compressed and uncompressed images, as disclosed in the '512 patent, the present inventors further determined that the model does not take into account the display parameters of the output device, such as the color primaries, the modulation transfer function, resolution (e.g., dpi), and tone scale. To overcome this limitation the present inventors determined that a model, such as a visual model of the human visual system, should be used as the basis of comparison between uncompressed and compressed images in the spatial domain.

Referring to FIG. 2, the system may include an input image 50 which is to be compressed using different Q-tables (or the same Q-table modified). The discrete cosine transform

coefficients 52 are calculated from the input image 50 (which may be in original form or modified by other techniques). Thresholding of the DCT coefficients may be performed, if desired. A set of quantization tables (Q-table) 54, 56, 58, and 60 are used to quantize the discrete cosine transform coefficients. Larger values in the Q-table typically result in a smaller

5   compressed file size, with larger compression artifacts. Similarly, smaller values in the Q-table typically result in a larger compressed file size, with smaller compression artifacts. The present inventors came to the realization that an "optimal" Q-table is not only dependent on the viewing condition, but is also dependent on the image itself. In the preferred embodiment, a set of four Q-tables may be used based upon the human visual contrast sensitivity function (CSF) using

10   different viewing distances (such as 11, 14, 17, and 19 inches). The resolution of the intended display, the modulation transfer function of the display, the display luminance characteristics of the display, the display color gamut of the display, the tone response curve of the display, may be taken into consideration when creating the Q-tables. For example, closer viewing distances will result in a flatter Q-table in the frequency domain, while farther viewing distances will yield a

15   steeper Q-table in which the higher order DCT coefficients are quantized more aggressively (with respect to the flatter Q-table).

       The resulting set of Q-tables include characteristics that account for one or more of the following properties, such as for example, the contrast sensitivity function of the human visual system, the viewing distances, resolution of the intended display, the display luminance

20   characteristics of the display, the display color gamut of the display, the tone response curve of the display, and the modulation transfer function of the display. In this manner, the Q-table is different than it would have been had one or more of these factors been omitted or added.

       The DCT coefficients, and hence the resulting image after encoding, are compressed to substantially the same compression ratio. The compression ratio, may be for example, each (or a

25   plurality of) resulting image is within 25% of the same size, within 10% of the same size, or within 5% of the same size. To achieve sufficient similarity in compression ratio the Q-table may be scaled and the image recompressed. Accordingly, the effect of each Q-table for compressing a particular image may be more effectively compared against the effect of other Q-tables if the resulting compressed image has a sufficiently similar compression ratio.

30   A model 62, 64, 66, and 68, such as a color visual difference model, may be used to

compare the differences between the original image 50 (or otherwise an image that has not been compressed) and an uncompressed version of the respective image after quantization using the respective Q-table 54, 56, 58, 60. A color visual difference model simulates the visual perception of the human eye. ~~Onee~~ One such model is X. Feng, J. Speigel, and A. Morimoto,

5   "Halftone image quality evaluation using color visual models", Proc. Of PICS 2002, p5-10, 2002, incorporated by reference herein. Such a model collapses to CIELAB for large patches of color. The model may be calibrated so that the threshold occurs at delta E 1.0, regardless of the frequency and background.

The model, based upon the viewing condition and display characteristics, may calculate

10   the visibility of the differences as a function of location in the image. The result may be a set of values, or for JPEG a single number, from the visual difference map. A variety of different metrics may be used, such as root mean square, median, 90th percentile, and 99th percentile. In the preferred embodiment, the 99th percentile is used and the threshold may be set to 1 delta E unit, which is approximately the visual detection threshold. The threshold may be adjusted

15   higher for applications where quality is not critical and storage is at a premium. The threshold may also be adjusted lower for applications that quality is critical, or the JPEG images may be viewed at a close distance.

Once the Q-table has been selected at block 70 based upon some criteria, the image 50 is compressed using a DCT, the selected Q-table, and encoding of the data, at block 72. The

20   resulting image is then reconstructed and compared against the image 50 using a model, such as the color visual difference model at block 74. If the resulting error metric E at block 76 is smaller than a low threshold (such as a threshold minus a tolerance value which may be within approximately 5% of the tolerance, if desired) then a scaling factor that scales the values in the Q-table is checked at block 78 to see if it is greater than a maximum value. The scaling factor

25   scales the Q-table in some manner and thus controls the amount of compression, which impacts the resulting image quality. If the scaling factor is not greater than a maximum value then the scaling factor is increased at block 80. Thus, block ~~80~~ 84 results from the case when the compression artifacts are below the visual threshold based upon some viewing condition and/or display. Therefore, the image may be compressed further to reduce the compressed image size

30   by increasing the scale factor. The selected Q-table is then re-scaled using the modified scaling

factor and the image 50 is then re-quantized using the modified Q-table. The quantized image is then reconstructed and evaluated against the image 50 using a model, such as the color visual difference model at block 74. The error metric is computed at block 76 and if the error is greater than a high threshold (such as a threshold plus a tolerance value) then the scaling factor that

5   scales the value in the Q-table is checked at block 82 to see if it is smaller than a minimum value. If the scaling factor is not less than the minimum value then the scaling factor is decreased at block 84. Thus, block 80 results from the case when the compression artifacts are above the visual threshold based upon some viewing condition and/or display. Therefore, the image may be compressed less to increase the compressed image size by decreasing the scale factor. The

10   selected Q-table is then re-scaled using the modified scaling factor and the image 50 is then re-quantized using the modified Q-table. The quantized image is then reconstructed and evaluated against the image 50 using the color visual difference model at block 74. The error metric is computed at block 76 and if the error is within tolerances a suitable Q-table and scaling factor (or otherwise modified Q-table) is selected. The image may be saved in a suitable file format, such

15   as JPEG or otherwise transmitted to a suitable destination at block 86.

In another embodiment, the Q-tables may be based upon other criteria. For example, the Q-tables may represent different power spectra in the image to be compressed. This aspects relates to masking, which in turn relates to supra-threshold perception (i.e., in supra-threshold perception, the contrast is higher and more masking typically occurs). As the level of overall

20   masking occurring in an image rises, the variation in sensitivities of the spatial frequency channels decreases. This implies a flatter Q-table will be appropriate for that image. In the case of image with very special characteristics, such as an application that has many images of striated texture (microscopic medical images), then the tables may reflect the oriented textures as well, and additional tables may be desirable.

25   Referring to FIG. 3, a graphical illustration is provided on one embodiment of a portion of the system. All illustrated an original image 100 is encoded 102, such as by a JPEG encoder. The encoded image is then reconstructed 104. The original image 100 and the reconstructed image 104 are modeled, such as by a color visual difference model 106. The model 106 provides a visual difference map of the image 108 from which an error metric 110 may be obtained.